February 1946

BIOMETRICS

Vol. 2 No. 1 BULLETIN

THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

FIELD SAMPLING FOR THE ESTIMATION OF WIREWORM POPULATIONS

D. J. FINNEY
Lecturer in the Design and Analysis of Scientific Experiment,
University of Oxford

1. Introduction

The statistician concerned with economic entomology is frequently asked to recommend a sampling technique for estimating the density of an insect population. The principles to be adopted in devising a suitable technique are now fairly well understood; their object is to obtain, by a method economical of time and effort, an estimate known to be sufficiently accurate for practical use, and, as a secondary consideration, some measure of that accuracy. (The intervention of the statistician has sometimes resulted in greater attention being given to standard errors than to the estimate of population!) The difficulties are usually more entomological than statistical, for the technique can only be decided in relation to the biology and ecology of the insect studied. Different species of the same genus, or even different stages in the life cycle of a single species may require very different sampling methods.

In this paper, attention will be restricted to a consideration of statistical problems arising in the development of a sampling technique for soil insects, and, as an illustration of these, the estimation of wireworm populations will be discussed in some detail.

2. Principles of Sampling

Even within a single field which provides conditions suitable for the life of a particular species of soil insect, the population density may vary widely from point to point. At any one point, the density may change with time, showing cyclic and secular trends as well as sudden changes induced by ploughing or other violent alterations of environment. Consequently, unless the sampling procedure involves counts of randomly or, at least, objectively selected portions of the population, the estimate may be seriously biased.

The sample used for estimating the mean population density in a field generally comprises a set of *sampling units*, small circles or squares of surface dug to a depth great enough to include all, or nearly all, insects of the species studied which are within the boundary of the unit. The estimate of population is then calculated from the mean number of insects per unit. The method of extracting insects from the soil, whether it be hand-sorting or a mechanical process, should be capable of discovering almost all of the species which are present in the sample, as inefficiency in extraction may vary from field to field, from day to day, or from worker to worker, thus ruining the comparability of results. Herein lies the chief criticism of any attempt to estimate populations from baiting records: the area within which the attraction of a bait will

operate cannot be exactly defined, and there can be no certainty that all insects in the vicinity will respond to the attraction.

When a sampling procedure is being developed, its precision must be examined; sampling units must then be taken at random, either from the whole field or in equal numbers from every one of a number of equal areas into which the field has been divided. The standard error of sampling is estimated from discrepancies between counts in units from the same part of the field. After the establishment of a sufficiently accurate technique, there is no need to calculate errors for every field sampled, but an occasional check should be made; indeed, the precision will be increased, though it can no longer be measured, by dividing the field into as many areas as sampling units and taking one unit from a random point in each area (5).

Clearly one important consideration is the size of the sampling unit. The labour of extracting the insects from the soil will be roughly proportional to the volume of soil, or, since the units are dug to a fixed depth, to the total surface area of the sample. Also, if samples have to be taken from field to laboratory for examination, this total area will be an important factor in the planning of transport. To dig many small units probably requires longer time than to dig a few large ones which cover the same total area, but, as a first approximation, the efficiencies of units of different sizes may be compared in terms of the total sample areas needed to give equal precision in the population estimates. For example, if units with surface areas u_1, u_2 have sampling variances per unit of v_1, v_2 respectively (measured on the same scale of population density), samples of total areas A_1, A_2 will lead to estimates of equal precision if

$$u_1 v_1 / A_1 = u_2 v_2 / A_2,$$

and the efficiency of the second unit relative to the first is then A_1/A_2.

If individual insects are distributed in the field entirely at random and independently of one another, the number per sampling unit will follow a Poisson distribution. The variance, σ^2, of the number per unit (of area u) is then equal to the mean number, μ; an easily verified consequence is that the variance of an estimate of population based on n units is inversely proportional to nu, the total area of the sample, so that all sizes of unit are equal in efficiency. Any departure from complete randomness will usually be in the direction of making σ^2 greater than μ— though possibly some form of competition might lead to a more even distribution than the Poisson, thus *reducing* σ^2— and the smallest units are then always the most efficient. The magnitude of the differences in efficiency required to be investigated, however, in order to assess how far the increased efficiency of a small unit compensates for the greater labour in the field.

3. *Sampling for Wireworms**

In recent years much attention has been given to problems connected with the estimation of wireworm populations, and the development of this work provides a good illustration of the use of sampling methods for soil insects. The first use of a soil sampling method for wireworms was probably that of Roebuck (4), but he did not investigate the precision of the method. Jones (1) reported the results of sampling about 50 fields, using sampling units of 1, ¼, or 1/16 square foot and 25, 50, or 100 units per field; he considered that his counts agreed with the Poisson law. On *a priori* grounds this conclusion seems unlikely to be correct, for the younger larvae might be expected not to have dispersed completely from points of oviposition and variations in soil characteristics and plant cover might also interfere with a completely random distribution. Cochran's examination of sampling results obtained by Ladell (2) verified that the error variance was generally greater than the mean count, and, as will be seen below, more extensive work has since confirmed this.

When the campaign for the ploughing of old grassland began in 1939, it was realized that the succeeding arable crops were likely

*The wireworm is the larval stage of the elaterid or 'click' beetle. The most important British species are **Agriotes lineatus, A. obscurus,** and **A. sputator,** though other genera are also found, notably **Athous** and **Cryptohypnus.** The larvae have a five year life cycle and will attack most British agricultural crops. The establishment of heavy infestations normally takes place in grassland; their long life cycle make wireworms a major pest of crops grown in the first few years after the ploughing of old grassland.

to suffer severe wireworm damage. The Advisory Entomologists of England and Wales, in the thirteen Provinces into which the country is divided for agricultural advisory work, therefore began a vast program of sampling, the Wireworm Survey of England and Wales (3). The purpose of this Survey was to discriminate between those fields in which careful choice of crop and cultivations would be needed in order to combat the wireworm danger and those in which low populations were unlikely to harm any but the most susceptible crops, and also to obtain further information about wireworm populations and their effect on crops. At first the standard technique was to take twenty 6 in. square units from a field of 10 to 20 acres, selecting a random pair of points within each one-tenth of the field area. Later, in order to make possible the sampling of a larger number of fields, the amount of soil per field had to be reduced. As a cylindrical core of 4 in. diameter (1/500,000 acre) had by then been shown to be more efficient than the 6 in. square, a standard of twenty 4 in. cores was used; rather more cores were taken from larger fields, approximately in proportion to the square root of the area, since the importance of accurate classification increased with the size of the field. The units were dug to a depth of 6 in., and twenty of the standard cores weighed about 1 cwt. In general the larvae were extracted from the soil by hand-sorting; only those exceeding 5 mm. in length were scored, as many of the smaller ones were likely to escape notice. In this way 25,000 fields were sampled before the end of 1942, and since then the total has probably increased to over 40,000. The figures discussed here refer only to larvae of species of the genus *Agriotes*, which formed by far the major part of the population in most of the fields.

After the accumulation of sufficient evidence on the sampling errors, the requirement of random selection of sampling points was dropped. Some Advisers then adopted systematic patterns of points, but no attempt at standardization was made and patterns were varied to suit the shapes of fields. A common scheme was to sample in two lines parallel to the sides of the field and as far apart as was each from the nearer side; ten sampling points were equally spaced along each line, or displaced laterally a fixed distance alternately to either side of these positions. These systematic patterns might lead to biased estimates, but the earlier work had shown no evidence that, for example, the population density at the edges of a field differed consistently from that at the centre. A slight bias was unimportant for the purposes for which the estimates were to be used, since all fields would be affected to a similar degree. No attempt was made to compare the merits of different systematic patterns, but those which are well spread over the whole field are unlikely to differ in precision markedly and consistently from field to field.

4. *Precision of the Estimates*

For the first fields sampled in the Wireworm Survey, the values of s^2/m were calculated (m is the mean count per sampling unit and s^2 the observed variance between sampling units); in general this quantity exceeded unity, and tended to increase with increasing m, thus indicating that the distribution of larvae within a field did not satisfy the Poisson law. An alternative test of the Poisson distribution, which was found more convenient, is to plot s/m (the coefficient of variation) against m; the ordinate should average $1/\sqrt{m}$ if the distribution law is the Poisson, but if the larvae are not distributed entirely at random the ordinate will be greater.

In order to estimate the average precision of sampling at different levels of population, and to assess the relative merits of different sampling units, the relationship between s and m had to be studied. It must be remembered that the problem of advising farmers on war-time management and cropping was urgent, so that neither time nor staff could be spared for a preliminary extensive investigation of sampling variation. The Poisson distribution was soon found to give too low a variance (5); an empirical relationship had therefore to be sought, and an intelligent guess at this had to be based on the results of the first few (about 100) fields sampled. Bartlett (6) has proposed that for data of this type the sampling variance may be related to the mean by the equation

$$s^2 = am + bm^2,$$

and Bliss (7) has proposed the alternative

$$s^2 = am^b;$$

a later examination has suggested that the second of these would satisfactorily fit the wireworm data, but, since s/m had been plotted against m in testing the Poisson hypothesis, the relationship was originally graduated by drawing a freehand curve on this diagram. Fields having the same value of m often showed wide variations in s, so that the exact form of the relationship could not be very accurately determined nor would it, if known, be a reliable guide to the sampling errors in individual fields. Nevertheless, as further data were collected, the original freehand curve was found to need only slight modification.

In the winter of 1940-41, 2272 fields which had been under grass in 1940 were sampled. From these the relationship between s and m could be estimated much more satisfactorily; for practical purposes, only a smoothed empirical curve was required and the true algebraic form was of no particular interest. Values of s/m corresponding to a given m were averaged and plotted against m, and again a freehand curve was drawn through the points; this curve (Fig. 1) shows the sampling errors to be substantially greater than for a Poisson distribution except at very low values of m. Below $m=0.8$ the ordinate is approximately $(1+0.28m)/\sqrt{m}$, above $m=1.0$ it is approximately $0.60+0.62/m$, and these formulae were used for interpolation.

TABLE 1

Efficiencies of Various Sampling Units Relative to the 4 in. Core

Population density		*Per cent efficiency of*		
No. per 20 4 in. cores	*1,000 per acre*	*6 in. square*	*2½ in. core*	*2 in. core*
1	25	92	100	100
2	50	75	101	101
4	100	69	102	102
12	300	65	106	107
20	500	62	111	113
40	1,000	59	120	126
80	2,000	58	131	139

Results for other years agreed closely with those shown in Fig. 1. Similar curves have been obtained for other sizes of sampling unit, and from them have been estimated the average relative efficiencies of the units at different levels of population (3). For example, for a population of one million per acre, Fig. 1 shows s/m for 4 in. cores to be 0.91, and the corresponding value for 2 in. cores is 1.62; the larger core has four times the area of the smaller, and the efficiency of the larger relative to the smaller is therefore $4x(0.91/1.62)^2 = 1.26$. At low populations, since the Poisson law is then almost satisfied, all units are of about equal efficiency, but, as Table 1 shows, at high populations the differences become important. The 4 in. core is much to be preferred to the 6 in. square, and the 2½ in. core has definite advantages over the 4 in., though two and a half times as many cores are required to give the same total area. The 2 in. core is very little superior to the 2½ in., and the difference scarcely repays the trouble of taking one and a half times as many cores. The ideal is possibly somewhere between 4 in. and 2½ in., the soil type being a factor affecting the choice, since small cores are difficult to use on heavy or stony soils. It may be noted that Ladell's results and, on re-examination, Jones's also are in agreement with these from the Wireworm Survey.

In the winter of 1941-42, efforts were made to obtain estimates of sampling variation from fields still in stubble after a cereal crop in 1941, as it was suspected that errors of sampling might be larger under those conditions on account of a tendency for the larvae to congregate in the rows. Results from 262 fields as is seen from the curve relating s/m to m, also shown in Fig. 1. At low populations, sampling in stubble is as accurate as in grass, but at populations of about 500,000 per acre 15 per cent more cores are needed in stubble to give the same accuracy as in grass; the necessary increase in cores rises to 50 per cent at populations of one million per acre, and may be as much as 100 per cent at two million.

When the curves in Fig. 1, and others similar, had been established, they were used to give the standard error and fiducial limits of any estimate of population (3,5), instead of assigning to each estimate an error calculated from that sample alone. Thus any real differences in sampling variation between fields of the same average population were ignored. In the main, advice on the cropping of a field had to be based on the estimated population, and,

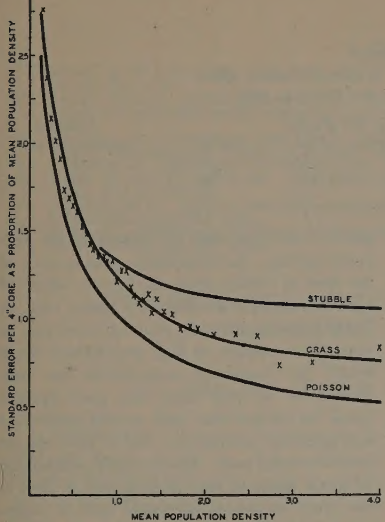

Fig. 1: Relationship between sampling error and mean population density for wireworm sampling by 4 in. core.
*means for 2272 fields under grass in 1940 and sampled in 1940-41.

providing that the technique and intensity of sampling were such as had previously been shown to give satisfactory accuracy on the average, the standard error of the individual estimates was not of great importance. When errors did not have to be calculated for every field, systematic sampling plans could be adopted with reasonable assurance that the errors would not be greater than predicted from Fig. 1, and all units from one field could be bulked if ease of transport and examination made this desirable. Bulking did, indeed, destroy the possibility of discovering fields which were heavily infested at one end, lightly at the other, but previous experience had shown that a difference of this kind sufficiently great as to justify different advice being given for the two parts occurred very rarely.

5. *Use of the Estimates*

The uses made of the estimates of population obtained in the Wireworm Survey were of two main types. Only the briefest of comments can be made on each type here, but a full account has been given elsewhere (3).

Firstly the estimates were used for the further study of wireworm populations *per se*, and of the factors influencing them. Geographical trends in population density and associations between soil characteristics and population were investigated, as also was the effect of arable cropping in reducing the population. Table 2 shows that even high populations are reduced to comparatively low average figures by two successive arable crops, and more extensive data from fields sampled in two consecutive years only have been used to construct curves which give, for any initial population, the expected population after one year of cropping. Comparison of results for individual fields with the expectations derived from these curves can then be used to indicate whether particular crops or cultivations are more or less effective than average in reducing the population.

TABLE 2

Mean Populations in Three Successive Winters of 104 Fields Ploughed from Grass in 1939-40 and thereafter under Arable Crops

	Mean Population (1,000 per acre)		
No. of fields	*1939-40*	*1940-41*	*1941-42*
52	160	170	100
23	440	360	170
15	810	440	230
14	1,380	710	270

Secondly the relationship between population density and the success of crops was studied. In general, detailed developmental study of a crop and accurate estimation of its yield were impossible, but information on plant density and yield samples were obtained from some fields. Efforts were made, however, to have as many crops as possible graded at harvest into one of three grades, and from these gradings tables such as Table 3 were compiled. These tables enabled the resistant and the susceptible of the more common crops to be discriminated very satisfactorily.

Application for entry as second-class matter is pending. The Biometrics Bulletin is published six times a year—in February, April, June, August, October and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N. W., Washington 6, D. C.

Membership dues in the American Statistical Association are $5.00 a year, of which $3.00 is for a year's subscription to the Quarterly Journal, fifty cents is for a year's subscription to the ASA Bulletin and members who pay $1.00 additional receive a year's subscription to the Biometrics Bulletin. Dues for Associate members of the Biometrics Section are $2.00 a year, of which $1.00 is for a year's subscription to the Biometrics Bulletin. Single copies of the Biometrics Bulletin are 60 cents each and annual subscriptions are $2.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N.W. Washington 6, D. C.

TABLE 3

Classification of 1,513 Crops of Spring Oats in 1941 on Fields under Grass in 1940

Wireworm population 1940-41 (1,000 per acre)	*Number of crops graded* Total	Satisfactory	Poor	Failed	*Percent unsatisfactory*
- 300	889	698	156	35	21
325 - 600	401	256	115	30	36
625 - 1,000	176	90	55	31	49
1,025 -	47	17	17	13	64

6. *A Warning*

For simplicity of presentation, the above account of the Wireworm Survey has been written as though the development of the Sampling Technique preceded and was entirely separate from its application. In fact, the urgent need of giving farmers practical assistance in combatting the wireworm required that the study of sampling, the investigation of relationships between wireworms, soils, management, and crops, and farm advisory work should proceed simultaneously. This had always to be borne in mind when attempting to use the records to give information of research value. The fields sampled were not randomly selected, but were chosen in response to requests for advice, and comparisons between mean populations for different regions or soil types may therefore be biased, though qualitatively if not quantitatively the main conclusions reached are likely to be correct. Again, records of fields where advice had been given in one year were for improving the next year's advice. An interesting consequence of the type of advice given is that the effect of high populations, as shown by tables such as Table 3, may apparently diminish from year to year. For example, at high populations wheat would be considered undesirable for poor farmers or on poor land, but good farmers would be advised that with careful cultivation and generous seeding they had a good chance of a successful crop. Consequently comparison of wheat crops on land at high and on land at low levels of population would tend to underestimate the true effect of wireworms in reducing the crop. As the Survey developed, such advice would be given more confidently and farmers would be increasingly willing to accept it, so that the apparent difference in results for wheat at extreme levels of population might be expected to diminish. Table 4 gives some evidence of the occurrence of this phenomenon.

TABLE 4

Percentages of Unsatisfactory Crops of Winter Wheat amongst All Crops Graded in 1941, 1942 and 1943

Wireworm population (1,000 per acre)	*Percentage unsatisfactory* 1941	1942	1943
- 300	28	22	21
325 - 600	42	31	25
625 - 1,000	50	46	32
1,025 -	61	48	40

7. *Summary*

The general problem of taking soil samples for estimating the population density of a soil insect has been considered. As an illustration, the technique used in the Wireworm Survey of England and Wales has been discussed. Though the figures analysed refer to one genus only, and exclude wireworms less than 5 mm. in length, the main findings are probably of wider application. Sample counts show the distribution of wireworms in a field to be not entirely random, so that small sampling units are more efficient, volume for volume, than large. The uses made of the estimates of wireworm population have been briefly indicated, and a caution about the uncritical drawing of conclusions from surveys of this type has been given.

The data on which this paper has been based were obtained by the Advisory Entomologists of England and Wales; most of the numerical results have already been published in references (3) and (5), but a small amount of previously unpublished material has been incorporated. I should like to express my thanks to all concerned for making these data available to me.

REFERENCES

(1) Jones, E. W. Practical field methods of sampling soil for wireworms. Jour. Agri. Res. 54:123-134. 1937.
(2) Ladell, W. R. S. Field experiments on the control of wireworms. With Appendix: The information supplied by the sampling results, by W. G. Cochran. Ann. Appl. Bio. 25:341-389. 1938.
(3) Ministry of Agriculture and Fisheries. Wireworm and food production: A wireworm survey of England & Wales (1939-1942). Bul. No. 128, pp. 62. 1944.
(4) Roebuck, A. Destruction of Wireworms. Jour. Min. Agri. 30:1047-1051. 1924.
(5) Yates, F. and D. J. Finney. Statistical problems in field sampling for wireworms. Ann. Appl. Bio. 29:156-167.
(6) Bartlett, M. S. Square root transformation in analysis of variance. Jour. Roy. Stat. Soc. Supplement 3:68-78. 1936.
(7) Bliss, C. I. Statistical problems in estimating populations of Japanese Beetle Larvae. Jour. Econ. Ent. 34:221-232. 1941.

THE ESTIMATION OF VARIANCE COMPONENTS IN ANALYSIS OF VARIANCE [1]

S. Lee Crump

Iowa State College, Statistical Laboratory

1. *Introduction.* Since its introduction in 1925 by R. A. Fisher (7), the analysis of variance has been used most widely to obtain tests of the significance of treatment effects. When he introduced this technique Fisher (7 § 40) indicated a further use of the analysis of variance. If an observed statistical variate, e.g. the plot yield of a varietal experiment, is assumed to be the sum of several separate effects (variety, block, etc. in the case of the varietal experiment), the variance of each effect will contribute to the total variance. The second use of the analysis of variance provides estimates of these several variance components. It is the purpose of this discussion to point out the hypotheses appropriate to the two uses of the analysis of variance and to explain its use to estimate variance components.

2. *The Hypotheses.* Although the hypotheses in each case are based on the same fundamental equation, they are essentially quite different. In order to illustrate, let us consider a randomized blocks experiment with treatments $A_1, A_2, \cdots, A_a$ arranged in blocks $B_1, B_2, \cdots, B_b$ and with n observations taken in each treatment in each block. If we denote by y_{hij} the j^{th} observation in the i^{th} block for the h^{th} treatment, we have for a fundamental equation

$$(1) \qquad y_{hij}=\mu+\alpha_h+\beta_i+\alpha\beta_{hi}+\zeta_{hij}. \quad \begin{array}{l} h=1,2,\cdots,a \\ i=1,2,\cdots,b \\ j=1,2,\cdots,n \end{array}$$

In equation (1) μ denotes the effect common to every observation, i.e. the general mean of y, α_h the effect common to every observation in A_h, β_i the effect common to every observation in B_i, $\alpha\beta_{hi}$ the effect common to every observation in both A_h and B_i, and ζ_{hij} the random effect peculiar to the hij^{th} observation.

In the first use of the analysis of variance, we have the following assumptions:

i. The ζ_{hij} are normally and independently distributed with mean zero and variance σ_ζ^2.
ii. The α_h, β_i, $\alpha\beta_{hi}$ and μ are parameters which remain fixed from sample to sample.

The problem here is to estimate the α_h, β_i, $\alpha\beta_{hi}$ and μ and to test the null hypothesis that any set of these parameters, $\alpha_1, \alpha_2, \cdots, \alpha_a$, say, are all equal to zero. The analysis of variance provides a solution to this problem.

In the second use of the analysis of variance, the assumptions are as follows:

i. The α_h, β_i, $\alpha\beta_{hi}$ and ζ_{hij} are all random variables independently distributed about means zero.
ii. The parameters in this case are the variances of α_h, β_i, $\alpha\beta_{hi}$ and ζ_{hij}, which we denote by σ_α^2, σ_β^2, $\sigma^2_{\alpha\beta}$ and σ_ζ^2 respectively, and μ.

Here the problem is to estimate σ_α^2, σ_β^2, $\sigma^2_{\alpha\beta}$

[1] Journal Paper No. J-1331 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project No. 890.

TABLE 1

The Analysis of Variance for a Randomized Blocks Experiment

Source of variation	Degrees of freedom	Mean square	Average value of the mean square
Treatments (A Classes)	$a-1$	M.S.$_A$	$\sigma_\zeta^2+n\sigma^2_{\alpha\beta}+nb\sigma_\alpha^2$
Blocks (B Classes)	$b-1$	M.S.$_B$	$\sigma_\zeta^2+n\sigma^2_{\alpha\beta}+na\sigma_\beta^2$
AB Interaction	$(a-1)(b-1)$	M.S.$_{AB}$	$\sigma_\zeta^2+n\sigma^2_{\alpha\beta}$
Error (within Subclasses)	$ab(n-1)$	M.S.$_Z$	σ_ζ^2

and σ_ζ^2. It should be noted that no assumption is made about the form of the distributions.

3. *The Average Values of the Mean Squares.* Table 1 shows the analysis of variance for our randomized blocks example. Under the assumptions of the second hypothesis above,it may be shown that the average values, in repeated samples, of the mean squares in column 3 are the expressions shown in the fourth column of this table.

A consideration of these expressions suggests that we may be able to formulate a general rule for determining the average values of any mean square from a multiple classification with equal subclass numbers. We will give such a rule. Consider an $ABCD\cdots$ classification with a A-classes, b B-classes, etc. and with n observations in each of the smallest subclasses. As in our preceding example, we write for a fundamental equation

$$y_{hijk\cdots}=\mu+\alpha_h+\beta_i+\cdots+\alpha\beta_{hi}+\alpha\gamma_{hj}+\cdots$$
$$+\alpha\beta\gamma_{hij}+\alpha\beta\delta_{hik}+\cdots+\alpha\beta\gamma\delta_{hijk\cdots}+\zeta_{hijk\cdots},$$
$$h=1,2,\cdots a$$
$$i=1,2,\cdots b$$
etc.

where α_h is common to all individuals in the h^{th} A class etc. As before let the variances of α_h, β_i, $\alpha\beta_{hi}$, $\cdots$ be σ_α^2, σ_β^2, $\sigma^2_{\alpha\beta}$, $\cdots$. In the analysis of variance for this classification, denote the mean squares in a manner corresponding to that of Table 1. With this notation the rule may be stated: "The average value of M.S.$_Z$ (mean square within subclasses) is σ_ζ^2. The average value of any other mean square is a linear combination of σ_ζ^2 and all other σ^2's whose subscripts contain every greek letter corresponding to the subscript of the mean square in question. The coefficient of σ_ζ^2 is unity. The coefficient of any other σ^2 is $\Delta=nabcd\cdot\cdots$ divided by the small roman letters corresponding to the greek subscript of the σ^2." Let us illustrate the rule with two mean squares from a fourfold $ABCD$ classification in which $a=2$, $b=3$, $c=3$, $d=4$ and $n=2$. We have $\Delta=2\cdot3\cdot3\cdot4\cdot2=144$. According to the rule the average value of M.S.$_{ABC}$ contains σ_ζ^2, $\sigma^2_{\alpha\beta\gamma}$ and $\sigma^2_{\alpha\beta\gamma\delta}$. Thus, we have

$$E(\text{M.S.}_{ABC})=\sigma_\zeta^2+\frac{\Delta}{abc}\sigma^2_{\alpha\beta\gamma}+\frac{\Delta}{abcd}\sigma^2_{\alpha\beta\gamma\delta}$$
$$=\sigma_\zeta^2+\frac{144}{2\cdot3\cdot3}\sigma^2_{\alpha\beta\gamma}+\frac{144}{2\cdot3\cdot3\cdot4}\sigma^2_{\alpha\beta\gamma\delta}$$
$$=\sigma_\zeta^2+8\sigma^2_{\alpha\beta\gamma}+2\sigma^2_{\alpha\beta\gamma\delta}$$

where $E(\ )$ denotes average value of the term in the parentheses. Similarly

$$E(\text{M.S.}_{AC})=\sigma_\zeta^2+24\sigma^2_{\alpha\gamma}+8\sigma^2_{\alpha\beta\gamma}$$
$$+6\sigma^2_{\alpha\gamma\delta}+2\sigma^2_{\alpha\beta\gamma\delta}.$$

This rule will apply to any multiple classification with equal subclass numbers. In the case that $n=1$ the rule is applied with σ_ζ^2 set equal to zero. In this case $\sigma^2_{\alpha\beta\gamma\delta\cdots}$ assumes the role of σ_ζ^2.

4. *Estimation of the Variance Components.* Returning to column 4 of Table 1 we see that

$$E(\text{M.S.}_{AB}-\text{M.S.}_Z)=n\sigma^2_{\alpha\beta},$$
$$(2)\qquad E(\text{M.S.}_B-\text{M.S.}_{AB})=na\sigma_\beta^2,$$
$$E(\text{M.S.}_A-\text{M.S.}_{AB})=nb\sigma_\alpha^2,$$

and hence that unbiased estimates of the σ^2's may be obtained from linear equations in the mean squares. If we let the "hat" ($\wedge$) denote "estimate of," we have for example

$$\hat\sigma^2_{\alpha\beta}=\frac{1}{n}(\text{M.S.}_{AB}-\text{M.S.}_Z),$$
$$\hat\sigma_\alpha^2=\frac{1}{nb}(\text{M.S.}_A-\text{M.S.}_{AB}).$$

In any multiple classification, when the average values of the mean squares have been written down, the linear combination of the mean squares which estimates a given σ^2 or variance component will be evident.

In most cases it is desirable to know the

variance of an estimated component. If each of the random elements in our fundamental equation follows a normal distribution, it may be shown that any analysis of variance mean square is distributed as $\frac{\chi^2\sigma_0^2}{f}$ where σ_0^2 is the average value of the mean square in question, f the corresponding degrees of freedom, and χ^2 follows the ordinary type III distribution with f degrees of freedom. Hence, since the variance of χ^2 is $2f$, the variance of any mean square is $\frac{2\sigma_0^4}{f}$. Further, it may be shown that the mean squares are independently distributed, and thus we may write the variance of any linear function of the mean squares. For example consider the first of equations (2). Let

$$E(\text{M.S.}_{AB}) = \sigma_\zeta^2 + n\sigma^2_{\alpha\beta} = \sigma_0^2,$$

and

$$E(\text{M.S.}_Z) = \sigma_\zeta^2 = \sigma_1^2.$$

Then since

$$\hat{\sigma}^2_{\alpha\beta} = \frac{1}{n}(\text{M.S.}_{AB} - \text{M.S.}_Z)$$

we may write

$$V(\hat{\sigma}^2_{\alpha\beta}) = \frac{2}{n^2}\left(\frac{\sigma_0^4}{f_0} + \frac{\sigma_1^4}{f_1}\right),$$

where $f_0 = (a-1)(b-1)$ and $f_1 = ab(n-1)$. Now σ_0^2 and σ_1^2 are unknown population values and we will obtain biased estimates of $V(\hat{\sigma}^2_{\alpha\beta})$ if we substitute M.S._{AB} and M.S._Z for σ_0^2 and σ_1^2 respectively. To correct for this bias, Daniels (4) gives the following formula to estimate V.

$$(3) \qquad \hat{V}(\hat{\sigma}^2_{\alpha\beta}) = \frac{2}{n^2}\left(\frac{\text{M.S.}^2_{AB}}{f_0+2} + \frac{\text{M.S.}^2_Z}{f_1+2}\right).$$

The variance of estimates of other components may be estimated in a similar manner.

In large samples the estimates are normally distributed under the condition of the preceding paragraph. Thus fiducial limits may be placed on the estimates when large samples are available.

5. *An Example.* We shall now illustrate the procedures described in the preceding sections with a numerical example. Our data are drawn from a series of genetic experiments on egg production and comprise the total number of eggs laid by each of 12 females from 25 races of *Drosophila melanogaster*, on the fourth day of laying, the whole experiment being carried out 4 times.[2] The analysis of variance for this set of data is shown in Table 2.

The rule for writing the average values of the mean squares given in section 3 may be verified for the last column of Table 2. Following equations (2) we obtain the following estimates of the variance components:

$$\hat{\sigma}_\zeta^2 = \text{M.S.}_Z = 231,$$

$$\hat{\sigma}^2_{\epsilon\rho} = \frac{1}{12}(\text{M.S.}_{ER} - \text{M.S.}_Z) = 19,$$

$$\hat{\sigma}_\rho^2 = \frac{1}{48}(\text{M.S.}_R - \text{M.S.}_{ER}) = 58, \text{ and}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{300}(\text{M.S.}_E - \text{M.S.}_{ER}) = 154.$$

Applying the principles used in obtaining equation (3), we have the following estimates of the variances of our estimated components:

$$\hat{V}(\hat{\sigma}_\zeta^2) = \frac{2(\text{M.S.}^2_Z)}{1102} = \frac{2(231)^2}{1102} = 97,$$

$$\hat{V}(\hat{\sigma}^2_{\epsilon\rho}) = \frac{2}{144}\left(\frac{\text{M.S.}^2_{ER}}{74} + \frac{\text{M.S.}^2_Z}{1102}\right) = 40,$$

$$\hat{V}(\hat{\sigma}_\rho^2) = \frac{2}{2304}\left(\frac{\text{M.S.}^2_R}{26} + \frac{\text{M.S.}^2_{ER}}{74}\right) = 354,$$

$$\hat{V}(\hat{\sigma}_\epsilon^2) = \frac{2}{90{,}000}\left(\frac{\text{M.S.}^2_E}{5} + \frac{\text{M.S.}^2_{ER}}{74}\right) = 9676.$$

[2] I am indebted to Dr. J. W. Gowen of Iowa State College for permission to use these data.

TABLE 2

Analysis of Variance of Total Egg Production of 12 Females (*D. melanogaster*) from 25 Races in 4 Experiments

Source of variation	Degrees of freedom	Mean square	Average value of the mean square
Experiments	3	$\text{M.S.}_E = 46{,}659$	$\sigma_\zeta^2 + 12\sigma^2_{\epsilon\rho} + 300\sigma_\epsilon^2$
Races	24	$\text{M.S.}_R = 3{,}243$	$\sigma_\zeta^2 + 12\sigma^2_{\epsilon\rho} + 4\sigma_\rho^2$
$E \times R$	72	$\text{M.S.}_{ER} = 459$	$\sigma_\zeta^2 + 12\sigma^2_{\epsilon\rho}$
Within Subclasses	1100	$\text{M.S.}_Z = 231$	σ_ζ^2

One may ask what sort of question may be answered with the help of the estimated variance components. We shall illustrate with one question. Suppose that it is desired to estimate the mean egg production of the i^{th} race on the fourth day of laying to some specified degree of accuracy. Let $\bar{x}_{\cdot i \cdot}$ be the mean egg production estimated from n females in each of e experiments. Then with the notation of section 2, we have

$$\bar{x}_{\cdot i \cdot}=\mu+\rho_i+\frac{\epsilon_i+\cdots+\epsilon_e}{e}+\frac{\epsilon\rho_{1i}+\cdots+\epsilon\rho_{ei}}{e}+\frac{\zeta_{1i1}+\cdots+\zeta_{ein}}{en}.$$

Hence we have for the variance of $\bar{x}_{\cdot i \cdot}$ about its average value, $\mu+\rho_i$

$$V(\bar{x}_{\cdot i \cdot})=\frac{1}{e}(\sigma_\epsilon^2+\sigma^2_{\epsilon\rho})+\frac{1}{en}(\sigma_\zeta^2)$$

and we may estimate this quantity by

$$\text{(4)}\qquad \hat{V}(\bar{x}_{\cdot i \cdot})=\frac{1}{e}(154+19)+\frac{1}{en}(231).$$

From equation (4) it is clear that, by increasing the number of experiments sufficiently, we may reduce $V(\bar{x}_{\cdot i \cdot})$ to any desired level. However, increasing the number of females per experiment indefinitely still leaves us with $V(\bar{x}_{\cdot i \cdot})=\frac{173}{e}$. In practice, the decision to increase n or e or both enough to reduce $V(\bar{x}_{\cdot i \cdot})$ to the desired level will depend on the relative costs of these alternatives.

6. *Estimation of Variance Components in More Complicated Cases.* When the classes have unequal numbers of individuals, it becomes extremely tedious to work out the average values of the mean squares. The principles, however, remain the same, involving straight forward algebra and elementary probability laws. Winsor and Clarke (14) have given the results for a one-fold classification into groups with unequal numbers. Consider data arranged in a groups with n_i observations in the i^{th} group ($i=1, 2, \cdots, a$). Let $N=\sum_{i=1}^{a} n_i$. In our previous notation M.S.$_A$ and M.S.$_Z$ denote the mean squares between and within groups respectively. Winsor and Clarke give the following average values for M.S.$_A$ and M.S.$_Z$:

$$E(\text{M.S.}_Z)=\sigma_\zeta^2 \quad\text{and}\quad E(\text{M.S.}_A)=\sigma_\zeta^2+n_0\sigma_\alpha^2,$$

where

$$n_0=\frac{1}{a-1}\left(N-\frac{1}{N}\sum_{i=1}^{a} n_i^2\right).$$

The average values of the mean squares for an extended "groups within groups" type of classification have been given by Ganguli (8), Hetzer, Dickerson and Zeller (9) and Finkner, Morgan and Monroe (6).

When the classification is of the $ABCD\cdots$ type with unequal subclass numbers, there are a number of methods of analysis available, e.g. the method of fitting constants, Yates (15), the method of weighted squares of means, Yates (15), and the method of expected subclass numbers, Snedecor (12). Each of these methods of analysis can give unbiased estimates of the variance components. At the present time, however, it is not known which method gives "best" estimates.

7. *Other Literature on the Estimation of Variance Components.* There are certain relationships between the results of the analysis of variance under the two hypotheses of section 2 which sometimes enable us to interpret tests of significance in terms of variance components. Discussions of such interpretations are given by Cochran (1), Crowther and Cochran (3), Wilm (13) and Yates and Cochran (17).

The use of the analysis of variance to estimate variance components has wide application in the selection of efficient sampling designs. Excellent discussions of these applications are found in Cochran (2), Yates and Zacopanay (16) and Youden and Mehlich (18).

Another field in which estimates of variance components are widely used is Genetics. Dickerson (5), Hetzer, Dickerson and Zeller (9), Lush and Molln (10) and Sprague (11) provide illustrations of applications in this field.

8. *A Note of Warning.* It must be remembered that in using the analysis of variance to estimate variance components, we have assumed the elements of the fundamental equation to be randomly selected from an infinite population. In an experiment

where three widths of spacing some crop are purposely selected for trial, it is not reasonable to regard these widths as random samples from all possible widths. On the other hand the blocks in a field experiment may sometimes quite reasonably be regarded as a random sample of all such blocks. In sampling production from say three machines in a factory, where these machines constitute all the machines which the factory has or is likely to have, it is more reasonable to regard these machines as the whole of a finite population than to consider them as random samples from some infinite population. If the factory owner is sampling production with a view to purchasing more machines of the same type, the three machines may be appropriately regarded as samples of the infinite population made up of all machines of the same type. Daniels (4) treats the case of finite populations. The assumptions must be considered carefully in any problem where it is desired to estimate variance components.

LITERATURE CITED

1. Cochran, W. G. Long term agricultural experiments. Jour. Roy. Stat. Soc., Supp. 6: 104–148. 1939.
2. Cochran, W. G. The use of the analysis of variance in enumeration by sampling. J. Am. Stat. Ass. 34: 492–510. 1939.
3. Crowther, F., and Cochran, W. G. Rotation experiments with cotton in the Sudan Gezira. J. Agric. Sci. 32: 390–405. 1942.
4. Daniels, H. E. The estimation of components of variance. Jour. Roy. Stat. Soc., Supp. 6: 186–197. 1939.
5. Dickerson, G. E. Experimental design for testing inbred lines of swine. J. An. Sci. 1: 326–341. 1942.
6. Finkner, A. L., Morgan, J. J., and Monroe, R. J. Methods of estimating farm employment from sample data in North Carolina. N. C. State Agric. Expt. Sta. Tech. Bull. No. 75. 1943.
7. Fisher, R. A. Statistical methods for research workers. Oliver and Boyd, London. Ed. 1. 1925.
8. Ganguli, M. A note on nested sampling. Sankhya 5: 449–452. 1941.
9. Hetzer, H. O., Dickerson, G. E., and Zeller, J. H. Heritability of type in Poland China Swine as evaluated by scoring. Jour. An. Sci. 3: 390–398. 1944.
10. Lush, J. L., and Molln, A. E. Litter size and weight as permanent characteristics of sows. U.S.D.A. Tech. Bull. No. 836. 1942.
11. Sprague, G. F., and Tatum, L. A. General vs. specific combining ability in single crosses of corn. J. Am. Soc. Agron. 34: 923–932. 1942.
12. Snedecor, G. W. The method of expected numbers for tables of multiple classification with disproportionate subclass numbers. J. Am. Stat. Ass. 29: 389–393. 1934.
13. Wilm, H. G. Notes on analysis of experiments replicated in time. Biom. Bull. 1: 16–20. 1945.
14. Winsor, C. P., and Clarke, G. L. A statistical study of variation in the catch of Plankton nets. Jour. Marine Res. 3: 1–34. 1940.
15. Yates, F. The analysis of multiple classifications with unequal numbers in the different classes. J. Am. Stat. Ass. 29: 51–66. 1934.
16. Yates, F., and Zacopanay, I. The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. J. Agric. Sci. 25: 545–577. 1935.
17. Yates, F., and Cochran, W. G. The analysis of groups of experiments. Jour. Agric. Sci. 28: 556–580. 1938.
18. Youden, W. J., and Mehlich, A. Selection of efficient methods for soil sampling. Contributions from Boyce Thompson Institute 9: 59–70. 1937.

ANALYSIS OF SCORES FROM SMELLING TESTS*

William Dowell Baten

Professor of Mathematics, Michigan State College, East Lansing, Michigan

Three small bottles containing 1%, 10% and 30% of rose extract, colored so that the strengths could not be detected by sight, were given to individuals, who were requested to arrange them in ascending order according to strength of aroma. To be able to determine whether or not those taking the test really showed ability to detect the different strengths of aroma, it is necessary to determine the number of correct arrangements which may be made by mere chance. A person, with no ability to smell, might by chance alone, arrange the bottles in the correct order. After it is known how many assortments on the average can be made by chance, then it will be possible to ascertain whether or not the judges were able to distinguish the various strengths of this material.

If the following is the correct arrangements of the bottles:

A	B	C
1%	10%	30%

then the following six arrangements, designated as (a), (b), (c), (d), (e) and (f), are possible:

Journal article No. 785 (N.S.) of the Michigan Agricultural Experiment Station.

TABLE I

	1%	10%	30%	Value
(a)	C	B	A	0
(b)	C	A	B	1
(c)	B	C	A	
(d)	A	C	B	2
(e)	B	A	C	
(f)	A	B	C	3

The order (a) means that the judge thought that the 1% solution was stronger than the 30% solution. This is the poorest judgment that he could make, although the second bottle is placed correctly. Order (b) means that the individual thought the strongest was the weakest, that the weakest was the next to the weakest, and that the next to the weakest was the strongest. This decision does not appear as bad a judgment as the above where the strongest and weakest solutions were interchanged. Order (c) indicates that the sampler thought that the weakest was the strongest concentration, that the next to the weakest was the weakest and that the 30% concentration was stronger than the 10%. Arrangement (d) shows that the smeller became confused on the two strongest aromas, but was able to distinguish the weakest from the two strongest. This person considered the 10% solution to be stronger than the 30% solution. Assortment (e) means that there was confusion about the correct ranks of the weakest and the next to the weakest. In the previous two orders, the tester became confused on the two adjacent strengths.

Let a value of 0 be given to (a), a value of 1 be given to a combination of (b) and (c), a value of 2 be given to a combination of (d) and (e) and a value of 3 be assigned to order (f), with respective probabilities 1/6, 2/6, 2/6 and 1/6. These weights are based upon the number of interchanges in an assortment necessary to secure the correct assortment.

Let the general case be considered, where p_0, p_1, p_2, and p_3 represent respectively the probabilities of four events 0, 1, 2, and 3 occurring (in our case arrangement (a), either (b) or (c), either (d) or (e) and arrangement (f)). The terms in the expansion of the multinomial

$$(p_0+p_1+p_2+p_3)^n = \sum \frac{n!}{n_0!n_1!n_2!n_3!} p_0^{n_0} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

(where the sum takes in all possible occurrences and where n_0, n_1, n_2, and n_3 represent respectively the number of times each event occurred) gives the respective probabilities of the various events; that is the first term, p_0^n, in this expansion is the probability of event 0 occurring n times; the second term, $np_0^{n-1}p_1$, is the probability that event 0 will occur $n-1$ times and event 1 will occur one time; the term, $\frac{n!}{i!j!s!g!}p_0^i p_1^j p_2^s p_3^g$, is the probability that event 0 will occur i times, event 1 will occur j times, event 2 will occur s times and event 3 will occur g times. Let the occurrence of event 0 correspond to a value of 0, the occurrence of event 1 correspond to a value 1, the occurrence of event 2 correspond to a value of 2 and the occurrence of event 3 correspond to a value of 3.

If $x = \sum_{i=0}^{3} in_i$, then x will range from 0 to $3n$.

The value $x=6$, is obtained when $n_0=n-6$, $n_1=6$, $n_2=n_3=0$, or when $n_0=n-5$, $n_1=4$, $n_2=1$, $n_3=0$, or when $n_0=n-4$, $n_1=2$, $n_2=2$, $n_3=0$, or when $n_0=n-3$, $n_1=0$, $n_2=3$, $n_3=0$, or when $n_0=n-4$, $n_1=3$, $n_2=0$, $n_3=1$, etc. Other values between 0 and $3n$ can be similarly analyzed. The following frequency distribution gives the various values of x corresponding to the above events, with their respective probabilities or frequencies.

X	*Probability*
0	p_0^n
1	$np_0^{n-1}p_1$
2	$\frac{n!}{(n-2)!2!}p_0^{n-2}p_1^2 + np_0^{n-1}p_2$
3	$\frac{n!}{(n-3)!3!}p_0^{n-3}p_1^3 + \frac{n!}{(n-2)!}p_0^{n-2}p_2p_1 + np_0^{n-1}p_3$
.	.
.	.
.	.
$3n$	p_3^n

It is necessary to determine the mean and standard deviation of this distribution. These can be found in the usual way or may be secured by taking the first and second

derivatives, at $z=0$, of the following expression

$$Y=(p_0e^{0z}+p_1e^{z}+p_2e^{2z}+p_3e^{3z})^n$$

The first derivative of Y at $z=0$ is found as follows:

$$Y'=n(p_0e^{0z}+p_1e^{z}+p_2e^{2z}+p_3e^{3z})^{n-1} \times(p_1e^{z}+2p_2e^{2z}+3p_3e^{3z}),$$

and

$$Y'_{z=0}=n(p_1+2p_2+3p_3)=\mu'_x.$$

The second derivative of y at $z=0$ is

$$\mu'_{2:x}=n(n-1)(p_1+2p_2+3p_3)^2 +n(p_1+4p_2+9p_3).$$

The second moment about the mean is

$$\sigma_x^2=n[(p_1+2p_2+3p_3)(p_0-p_2-2p_3) +2p_2+6p_3].$$

The standard deviation of the above distribution is

$$(2)\quad \sigma_x=\sqrt{n[(p_1+2p_2+3p_3)(p_0-p_2-2p_3) +2p_2+6p_3]}$$

Two hundred people took the smelling test described above with the results given in Table 1.

TABLE 1.

Number of people arranging the bottles with arrangements (a), (b) and (c), (d) and (e) and (f), where i represents the corresponding values.

i	Frequency	Percentage
0	18	.090
1	39	.195
2	73	.365
3	70	.350

The total of the scores or the value of x is 395; this total was obtained as follows:

$$(0\times18)+(1\times39)+(2\times73)+(3\times70)=395.$$

Is this total of 395 significantly different from the total score which 200 people would have made, on the average, if they had no ability to smell and had arranged the bottles by mere guessing? If chance alone were used, the average total according to (1) is

$$\mu'_x=n(p_1+2p_2+3p_3)=200(2/6+4/6 +3/6)=300,$$

since $p_0=1/6$, $p_1=2/6$, $p_2=2/6$ and $p_3=1/6$. According to (2), the standard deviation is

$$\sigma_x=\sqrt{200\left[\left(\frac{9}{6}\right)\left(-\frac{3}{6}\right)+\frac{10}{6}\right]}=13.5$$

The t-value will show whether or not the observed total of 395 is significantly different from the average total of 300, which was due to chance. The t-value with an infinite number of degrees of freedom is

$$t=\frac{395-300}{13.5}=7.0,$$

which indicates that the judges were not arranging the bottles by chance. The testers were evidently using their sense of smell in arranging the bottles in the required order.

Attempts at removing the effects of the last aroma.

Some of the judges said that after they had smelled the first bottle their olfactory nerves were saturated with the aroma, so that they were not able to detect stronger or weaker strengths afterwards. To overcome this difficulty several materials were used between the bottles of rose extract.

The same test was given to 73 mature people with the condition that a bottle of peppermint was to be held to the nose before smelling the second bottle of rose extract and also before smelling the third bottle of rose extract. The object of this test was to determine whether or not the aroma from the peppermint removed the aroma of the rose extract from the olfactory nerves and enabled the judge to better distinguish the strength of the next bottle of rose extract. In some taste experiments the tasters eat a cracker, take a sip of water, or eat a slice of an apple to remove the effects of the last sample from the taste buds. Table 1 contains the expected percentages of the individuals that on the average arrange the bottles as (a), (b) and (c), (d) and (e), and (f). These percentages were obtained from the 200 people taking the test and constituted all the information concerning the ability of people to assort the bottles according to strengths of aroma. If the above percentages are the true ones, and they are used, then the mean and standard deviation of the values from 0 to 219, whose frequencies are the respective terms of the expansion of the multinomial

$$(.090+.195+.365+.350)^{73}$$

are equal respectively, to

$$\mu'x=144 \text{ and } \sigma_x=8.2$$

The second column of Table 2 contains the results of the test when peppermint was used between the bottles of rose extract. The total score or value found from these values was as follows:

$(0\times9)+(1\times13)+(2\times30)+(3\times21)=136$

How does this total of 136 compare with the expected total of 144 when no peppermint was employed?
The t-value is

$$t = \frac{136-144}{8.2} = -0.98$$

which indicates that the peppermint did not better arrange the three strengths of rose extract. Some of the judges thought that these materials did aid them, others thought that they hindered them, and some could not detect any assistance or any hindrance. The expected frequencies in Table 2 indicate how near they are to the observed frequencies; clearly the aromas used between the vials did not help or hinder the judges in assorting the three strengths of aroma.

TABLE II

Table 2. Frequently distributions of the values assigned to arrangements (a), (b) and (c), (d) and (e), and (f), where i represents the corresponding values

	Materials used between bottles of rose extrace							
	Peppermint		*Turpentine*		*Vanilla*		*Vinegar*	
i	*Observed freq.*	*Exp. freq.*	*Observed freq.*	*Exp. freq.*	*Observed freq.*	*Exp. freq.*	*Observed freq.*	*Exp. freq.*
0	9	6.6	6	6.1	5	6.8	12	5.8
1	13	14.2	11	13.3	13	14.6	10	12.5
2	30	26.6	28	24.8	26	27.4	23	23.3
3	21	25.6	23	23.8	31	26.3	19	22.3
No. of people N	73	73.0	68	68.0	75	75.0	64	64.0
Observed total	136		136		158		113	
Expected total	144		134		148		126	

affect the decisions pertaining to assorting the rose extract solutions.

A bottle of turpentine was used between the bottles of rose extract as was done with peppermint to determine whether or not a whiff of turpentine before going to the next bottle in the test removed the aroma in the nose. Vanilla extract and vinegar were used similarly. The results from these tests also are contained in Table 2 together with the total scores and the expected scores.

On examining the respective t-values, it is found that there were no significant differences between the actual totals and the expected totals, which show that the turpentine, vanilla extract and vinegar did not help the judges to assort the rose extract solutions.

It is interesting to note, in the experiment in which 200 people were used, that the total score or value of 395 was far from a perfect score of 600. Thirty-five per cent were able to assort all of the bottles correctly; 28.5% assorted them as either (a), (b) or (c). This means that these strengths of 1 per cent, 10 per cent and 30 per cent could not be detected accurately by the majority of the people and that people differ considerably in their ability to distinguish aromas.

If a food laboratory desires to select a panel of judges for detecting aromas of certain foods it would be advisable to carry out a similar test and select those who were able to arrange the material correctly.

QUERIES

QUERY: I have a problem in mind which I would like to have answered if possible.

Source of variation	Degrees of freedom	Mean Square
Total	99	
Blocks	9	66.667
Treatments	9	33.333
Error	81	1.234

In the above table of analysis of variance both "Blocks" and "Treatments" are significant at the 1% level. However, the value of F for Blocks is twice as great as the value of F for Treatments. Does this mean that the variance accounted for by Blocks is twice that accounted for by Treatments, or might the variance accounted for by Blocks and

that accounted for by Treatments be approximately the same for both Blocks and Treatments since both are significant at the 1% level?

I am concerned about the interpretation of the relative importance of the two significant variables.

ANSWER: Following the ideas developed in the paper by this writer in the present issue of the Biometrics Bulletin and taking the assumptions outlined there as fulfilled, we find for estimates of σ_β^2, σ_τ^2, and σ_ϵ^2 the following:

$$\hat{\sigma}_\beta^2 = 6.543$$
$$\hat{\sigma}_\tau^2 = 3.210 \quad \text{and}$$
$$\hat{\sigma}_\epsilon^2 = 1.234$$

where the subscripts β, τ, and ϵ, refer to the blocks, treatments and error components of variance respectively. Hence, in your example with its small error mean square, the estimate of the variance ascribable to blocks is slightly more than twice that of the variance ascribable to treatments. However, these estimates are subject to a large sampling error. By approximate methods we may place fiducial limits on $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_\tau^2$. The 5% fiducial limits in this example are 3.423–17.919 for $\hat{\sigma}_\beta^2$ and 1.650–8.900 for $\hat{\sigma}_\tau^2$. Any interpretation of the relative importance of the two main effects should take into account this large sampling variation.

S. Lee Crump

QUERY It is usual to plot the independent variable on the X axis and the dependent variable along the Y axis, but Ezekiel, 1930, pp. 129 and 131, reverses the procedure. Since I was trying out a linear and a curvilinear regression, that change bothered me. Is there any rule about this?

ANSWER It is conventional, as you say, to plot the independent variable along the horizontal axis. Ezekiel apparently does this, the amount of water applied being considered the independent or controlled variate with cotton yield as dependent.

Perhaps you are thinking that at any particular stage of growth the size attained by the plant determines the amount of its water requirement. If so, then you would doubtless plot the size of the plants as X along the horizontal axis. Ezekiel's irrigation problem required a different attitude toward the cause and effect relation of the variates.

George W. Snedecor

QUERY In a covariance problem, the question arose as to how to adjust the treatment means for differences in their values of the independent variate. Will you explain this for a randomized blocks experiment?

ANSWER Ordinarily the treatment means are adjusted by use of the error regression coefficient. This results in the set of adjusted means, differences among which are tested by the usual procedure. The following data are needed:

1. The error regression coefficient, $b = Sxy/Sx^2$.
2. The deviation of the treatment mean of the independent variate from the experiment mean. Denote this by x.
3. The treatment mean of the dependent variate designated as Y. The adjusted value is then Y—bx.

In Fisher's example 46.1, the error row in the covariance table has $Sx^2 = 567.5$ and $Sxy = 654.25$; hence, $b = 654.25/567.5 = 1.1529$.

If the mean of all values of X is 100 while the mean X for a certain treatment is 92.25, then $x = 92.25 - 100 = -7.75$.

The mean Y for the same treatment is 82.25. Finally, the adjusted mean Y is $82.25 - (1.1529)(-7.75) = 91.18$.

George W. Snedecor

QUERY In the first query of the August, 1945, *Biometrics Bulletin*, page 55, the questioner ends his problem with the following question: "What . . . is the probability that the . . . samples are not merely samples from the same population?" Obviously this question cannot be answered without *a priori* knowledge of the probabilities of the populations.

At first thought one might regard the error in the question as merely a slip of the tongue or poor phraseology. However, recent experience with statisticians and experimentalists trained in various colleges has convinced me

that it reflects more often than not a serious error in thought. They persistently are found writing or saying that the small probability derived from a significant test is the probability of there being a true difference in the populations sampled!

In fact one need not rely upon inexperienced statisticians as a source of such errors. If memory serves me right, the same error is noticeable in some textbooks.

The answer to the query gave the probability that two independent random samples from the *same* population could produce two interactions differing (as measured by their ratio) more than the ratio obtained from the two observed samples. Should you not have rephrased the question so as to make it clear what question you were answering?

ANSWER Yes; and thank you for calling attention to the discrepancy.

Other readers have raised the question as to whether I really got at the root of the original querist's difficulty. Of the seven degrees of freedom in the 2×2×2 table, only one was discussed, leaving three main effects and three first order interactions unnoticed. So far as I have been able to see, these other six effects are not relevant to the query.

George W. Snedecor

QUERY In a cultivation experiment on sugar beet, three treatments were tried in seven randomized blocks. The analysis of variance and the treatment means are as follows:

Source of variation	*Degrees of freedom*	*Mean square*
Treatment	2	1.534
Block	6	3.409
Error	12	0.232
	Mean yields in tons per acre	
Treatment	1	15.88
	2	15.00
	3	15.18

For treatments, $F = 6.61$, which does not reach the significant value, 6.93. On the contrary, the ratio of the range of means to $s_{\bar{x}}$, $0.88\sqrt{0.232/7} = 4.84$, is beyond its 1% value, 4.10. This was calculated as indicated in the April *Bulletin*.

It is my understanding that no significant differences exist if the calculated F value is less than that found in the table. Thus, from the F test, I conclude that there are no differences, while from the test of range the opposite is true. I have heard several explanations of this discrepancy but would appreciate having your comments.

ANSWER From my point of view, there is no discrepancy, because the probabilities turned up by the two tests are practically the same. If one probability were 0.1 and the other 0.01, I should rely upon the probability of F as being the more dependable; but I should also go over the whole experiment again to learn why the two indications were so different.

Perhaps your difficulty lies in a somewhat arbitrary interpretation of the test. You imply that if P is a bit larger than 0.01 there are no differences, while if P is just a little smaller than that value, then suddenly there are differences. This is not a realistic attitude toward a test of significance. There is no sharp break in the scale of probability but rather a continuous flow. The conventional 0.05 and 0.01 points are like mileposts on a road: you do not jump from 6 miles to 5 as you pass the post. Let us review the steps taken in a test of significance.

First, one sets up a null hypothesis in conformity with the objective of his experiment—in your case, the hypothesis is that the three treatment means are drawn from a common population; that is, that the treatments do not affect yield. Next, one calculates F for the experimental yields. Third, one observes in a table the probability of a larger F in random sampling from the hypothetical population. In your experiment, the probability of a larger F is about 0.01; that is, one would expect to get a larger F about once per hundred trials if the treatments were wholly ineffective. Finally, the experimenter makes a decision. He may decide that, despite the improbability of his sample, he will not reject the hypothesis set up: he may not be convinced that treatment 1 is superior in the population from which the experimental sample is drawn. Usually, however, with P=0.01, he would reject the null hypothesis, concluding that the treatments do affect yield, and that in the sampled popula-

tion treatment 1 is superior to the other two.

It should be observed that the experimenter might make either of these decisions if the experiment had resulted in other levels of probability. It is customary to decide in advance the level that will lead to rejection, but it should be clear that the statistical evidence is only part (and sometimes a small part) of the entire information upon which the decision of the experimenter must be based. There is no value of P at which it can be said with certainty that differences do or do not exist in the population.

George W. Snedecor

QUERY In a study of the ability of examiners to predict training success, a two by two table was secured for each examiner. Tables for eight examiners are given below. The problem is to test the hypothesis that the examiners do not differ among themselves with respect to their ability to predict training success.

Examiner	*Actual Outcome*	*Examiner's Prediction Pass*	*Prediction Fail*
1	Pass	6	2
	Fail	16	20
2	Pass	11	2
	Fail	19	5
3	Pass	7	2
	Fail	10	14
4	Pass	7	1
	Fail	8	2
5	Pass	3	1
	Fail	7	11
6	Pass	8	0
	Fail	4	12
7	Pass	5	1
	Fail	9	5
8	Pass	5	0
	Fail	9	5

ANSWER Ordinarily to test a 2×2×r table for homogeneity with respect to the r-fold classification, one would employ a χ^2 test for which 2r—2 degrees of freedom are available. In this particular example, however, it seems worthwhile to make two tests each involving r—1 degrees of freedom as will be explained below.

The sums for the above 2×2×8 table over the 8-fold classification are:

Actual Outcome	*Examiners' Prediction Pass*	*Fail*	*Totals*
Pass	52	9	61
Fail	82	74	156

Thus the examiners correctly predicted that 52 of the 61 passing students would pass, and that 74 of the 156 failing students would fail. They obviously did far better with the passing students than with the failing students. The reason for this is easy to surmise—the examiners wished to avoid predicting that a passing student would fail and in order to avoid doing so, they were willing to include a number of doubtful cases in their passing category. Thus the examiners predicted failure only when they felt fairly certain that the student would fail; all other students were predicted to pass.

The 9 passing students who were predicted to fail were, therefore, gross errors on the part of the examiners; while the 82 failing students who were predicted to pass merely represent a margin of safety used by the examiners in their endeavor to include all passing students in their passing prediction.

Two comparisons may be made between the examiners. A comparison using only the passing students will indicate whether the examiners differ in their proportions of gross errors having selected arbitrarily their own margins of safety against gross errors. The χ^2 test of homogeneity on the 2×8 table involving the 61 passing students gives $\chi^2 = 4.7$ with 7 degrees of freedom, which corresponds to a probability of about .80.

A comparison using only the failing students will indicate whether the examiners differ in their margins of safety. Here $\chi^2 = 19.7$ with 7 degrees of freedom and the probability is less than .01. Thus while all examiners miss a small proportion (about 15%) of the passing students, some examiners must pass significantly more failing students than others to achieve this small proportion.

If we may assume that it is easy to predict correctly the outcome for very capable students and very dull students, then the examiners should be judged on their treatment of borderline cases. The comparison using only the failing students enables one to judge the examiners on their treatment of borderline cases, but this is a circumstance of the data,

and is possible only because the first χ^2 test did not reject the null hypothesis (that there was no difference between examiners with regard to their commission of gross errors).

If, as might well have happened, some examiners had used a small margin of safety and made few errors in judging the failing students while making several errors in judging passing students; then it would have been appropriate to compare the examiners by the ordinary χ^2 test of homogeneity for 2×2×r tables. In the example at hand, this χ^2 would simply be the sum of the two χ^2's obtained above, $\chi^2=24.4$ with 14 degrees of freedom, which corresponds to a probability of about .04.

A. M. Mood

QUERY: We are planning to test the adaptability of some 28 grasses and legumes in this area when grown on 4 different major soil types. According to our plans we will have one nursery on each of the 4 different soil types which will be rather widely distributed in this area. We feel that it would facilitate our work somewhat if we would use the same random distribution in each of 4 locations. Why would it not be desirable to use the same random plan in each of the 4 locations?

ANSWER: The probability values given in the standard tables for tests of significance were calculated from certain assumptions about the nature of experimental data. One of these assumptions is that the experimental errors which affect different units are independent of one another. This assumption may not be justified. In fact, in field experiments, the yields of neighboring plots are usually found to be positively correlated; which means that the experimental errors on neighboring plots also are correlated. Randomization effectively avoids bias from this source and permits the data to be treated as if errors were independent.

Experiments have the common characteristic that, when they are repeated, the observed effects of the treatments vary from trial to trial. This variation introduces a degree of uncertainty into the interpretation of the results. The results are said to be subject to experimental error. Whatever the source of the errors, replication of the experiment decreases the error associated with the average effect of any treatment, provided that certain precautions are taken. These precautions must ensure that one treatment is no more likely to be favored in any replication than another, so that errors affecting any treatment tend to cancel out on the average as the number of replications is increased. One essential safeguard is that the treatments be assigned to the experimental units at random. That is, the purpose of randomization is to avoid any systematic bias arising from the differences among the experimental units. Every treatment receives an equal chance of being assigned to a favorable or unfavorable set of units. If treatment 1 and 4 appear adjacent in two replications in one location, and then the same randomization is used in four locations, you may have a biased pooled estimate of error for testing differences between these two treatments, although the estimate of the mean difference might be very accurate.

The purpose of randomization is to guarantee the validity of the test of significance, this test being based on an estimate of error secured by replication. Therefore, it is not desirable to use the same random plan in each of 4 locations.

Gertrude M. Cox

ANNUAL MEETING OF THE BIOMETRICS SECTION

The annual business meeting of the Section was held at the Statler Hotel in Cleveland at 12:30 on January 25, 1946, with 28 members and 13 guests present. Chairman Bliss reported on activities and development since the last previous annual meeting (September, 1944). This report appeared in the December issue of the Biometrics Bulletin.

The nominating committee, consisting of Prof. W. G. Cochran and Miss Besse Day, reported their nominations. In the absence of additional nominations, the committee's nominees were elected unanimously. The officers of the Section for the new year are: Chairman, D. B. DeLury; Secretary, H. W. Norton; Section Committee, E. J. deBeer, A. E. Brandt,

J. W. Fertig, J. G. Osborne, and J. W. Tukey; Editorial Committee, Gertrude Cox, Editor; C. I. Bliss, W. G. Cochran, Churchill Eisenhart, F. R. Immer, H. W. Norton, G. W. Snedecor, and C. P. Winsor.

Chairman Bliss outlined the plan for reorganization of the American Statistical Association and opened discussion of similar plans for the Section. President Shewhart of the Association commented briefly, suggesting that the Section move in the direction of autonomy, the main advantage being independent control of its activities, especially in technical matters, but that it retain close ties with the American Statistical Association as the central organization of all statistical groups. Secretary Norton, speaking for A. E. Brandt, Chairman of the Constitution Committee, presented a draft Constitution for the Biometrics Society, which it is proposed that the Section become, and indicated some of the questions which must be decided before a satisfactory constitution can be written. The discussion which followed centered around the advantages and disadvantages of organizing as a Society rather than as a Section. It was emphasized that discussion of this question and of the constitution should be completed in 1946, so that the question can be voted upon at the next annual meeting. Following approval of a resolution of commendation and appreciation of the work of the retiring chairman, the meeting was adjourned.

NEWS AND NOTES

The Executive Committee of the Animal Vitamin Research Council (AVRC) met in New York City on December 15, 1945. The afternoon session was open to the general membership and included a report by C. I. BLISS, Chairman of the Statistical Committee on the analysis of data from a collaborative study by the AVRC of suggested rations of the A. O. A. C. chick assay for Vitamin D. . . E. L. LeCLERG, pathologist, Bureau of Plant Industry, has been stationed at Baton Rouge, La. for several years. He is leaving the U. S. D. A. this spring to work in the Executive office of the President in the Bureau of Budget as a budget examiner. His duties will be those of a liaison officer between the research agencies of the U. S. D. A. and the Bureau of Budget. D. M. SEATH and Mr. LeClerg have been giving a course in statistics to a group of experiment station workers in the fields of nutrition and animal sciences. . . WILLIAM G. MADOW left the last of January for Sao Paulo, Brazil, to serve as Visiting Professor of Statistics at the University of Sao Paulo for the full academic year which begins on March 15. He expects to return to the United States early in January, 1947. . . S. C. SALMON and VICTOR R. BOSWELL have gone to Tokyo and other points in Japanese occupied territory to help with problems relating to food production in the area. Mr. Salmon will deal with cereal and field crops while Mr. Boswell will be helping with fruit and vegetable crop problems. A report will be expected when they return to the U. S. . . ROBERT PENQUITE is now with the Poultry Department at Iowa State College. . . RALPH E. COMSTOCK, Plant Science Statistician, Institute of Statistics, recently spent ten days at Piedras, Puerto Rico, to review the designs and objectives of the Animal Husbandry research projects of the Puerto Rico Agricultural Experiment Station. . . CARL F. KOSSACK spent the first part of the war at the University of Oregon teaching under the ASTP program. During the last year he was in Washington, D. C., as a consultant with the Office of Field Service of the O.S.R.D. and was assigned to work for the Air Forces. He has returned to the University of Oregon. . . ERIC C. WOOD, Virol Ltd. (food specialists), Hanger Lane, Ealing, England, writes, "The Society of Public Analysts and Other Analytical Chemists agreed in principle some time ago to the formation within itself of Groups for the study and furtherance of knowledge of such specialized branches of analysis as might well be of interest only to a certain proportion of its members. A Biological Methods Group has recently been formed, having as its field the use of biological methods in chemical analyses." This group held its inaugural meeting on October 17, 1945, at which time it was heartily agreed that their meetings would have a special interest to statisticians, since their work was so largely complementary to that of

the bio-assayist. . . ALBERT LEE SCHRADER, Dept. of Horticulture, University of Maryland, College Park, Md., taught horticulture at Shrivenham American University for awhile, then went to Amsterdam to discuss research activities in Agricultural Science, Engineering and Commerce. He states, "I hope to get a number of new ideas of interest from the Dutch people." During his stay in England, Mr. Schrader had the opportunity to visit several research stations. . . The work of D. BOYD SHANK, University of Arkansas, Fayetteville, consists of breeding and variety testing on corn and cotton. It is interesting to note that plant breeders now speak of the incomplete block designs as "conventional". . . . WAYNE F. FREEMAN, Bureau of Plant Industry, Experiment Station, Tifton, Ga., says the same thing: "My designs of experiments do not vary much from the now conventional lattices, triple lattices and lattice squares." Mr. Freeman finished his Ph.D. at Illinois last spring. . . We were glad to hear from C. H. GOULDEN, Dominion Rust Research Laboratory, Winnipeg, Manitoba. . . . It has been an effort to see that some of the biological statisticians become subscribers to this *Bulletin*. That and secret war work has made it difficult to secure articles. Send us one, but remember, we want a discussion of the use of a statistical technique to biological research data. Inclusion of a numerical example is urged.

Material for the BULLETIN should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C., material for Queries should go to "uQeries," Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.